



Review



Cite this article: Engeser M, Ajith S, Duymaz I, Wang G, Foxwell MJ, Cichy RM, Pitcher D, Kaiser D. 2025 Characterizing internal models of the visual environment. *Proc. R. Soc. B* **292**: 0602. <https://doi.org/10.1098/rspb.2025.0602>

Received: 3 March 2025

Accepted: 25 July 2025

Subject Category:

Neuroscience and cognition

Subject Areas:

cognition, neuroscience, behaviour

Keywords:

visual perception, scene representation, predictive processing, internal models, individual differences, drawings

Author for correspondence:

Daniel Kaiser

e-mail: danielkaiser.net@gmail.com

[†]These authors contributed equally to the study.

Characterizing internal models of the visual environment

Micha Engeser^{1,2,†}, Susan Ajith^{1,†}, Ilker Duymaz¹, Gongting Wang^{1,3}, Matthew J Foxwell⁴, Radoslaw M Cichy³, David Pitcher⁴ and Daniel Kaiser^{1,2,5}

¹Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, Giessen, HE, Germany

²Center for Mind, Brain and Behavior, Universities of Giessen, Marburg, and Darmstadt, Marburg, HE, Germany

³Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

⁴Department of Psychology, University of York, York, UK

⁵Cluster of Excellence "The Adaptive Mind", Universities of Giessen, Marburg, and Darmstadt, Giessen, HE, Germany

GW, 0000-0002-2364-0639; DK, 0000-0002-9007-3160

Despite the complexity of real-world environments, natural vision is seamlessly efficient. To explain this efficiency, researchers often use predictive processing frameworks, in which perceptual efficiency is determined by the match between the visual input and internal models of what the world should look like. In scene vision, predictions derived from our internal models of a scene should play a particularly important role, given the highly reliable statistical structure of our environment. Despite their importance for scene perception, we still do not fully understand what is contained in our internal models of the environment. Here, we highlight that the current literature disproportionately focuses on an experimental approach that tries to infer the contents of internal models from arbitrary, experimenter-driven manipulations in stimulus characteristics. To make progress, additional participant-driven approaches are needed, focusing on participants' descriptions of what constitutes a typical scene. We discuss how recent studies on memory and perception used methods like line drawings to characterize internal representations in unconstrained ways and on the level of individual participants. These emerging methods show that it is now time to also study natural scene perception from a different angle—starting with a characterization of an individual's expectations about the world.

1. Natural vision and internal models of the world

Perceptual efficiency is often understood through the lens of predictive processing [1,2]. In this framework, visual inputs are routinely compared against internal models, which are based on our expectations of what the world should look like. In the processing of natural environments, internal models should play a particularly helpful role [3,4]: natural scenes are reliably structured, with a global structure that is stable across instances of a category and objects placed in statistically predictable locations [5–10]. The reliable structure of natural scenes should give rise to rich internal models that capture what a specific scene (e.g. a kitchen) should typically look like.

The study of vision as an inverse inference problem has its origins in Helmholtz's idea of perception [11]. Within this framework, the perceptual system uses prior knowledge about the world, obtained through experience, to infer the causes of proximal stimulus patterns. In this view, internal models of the world, which contain this prior knowledge, are thus critical determinants for further efficient natural perception. This was later highlighted by

schema theory, which postulated that inputs are referenced against internal models (schemata) that reflect the structure of the world (e.g. the likely object arrangements found in a scene; [12,13]). This concept has influenced early research on scene perception [14,15] and memory [16,17]. In contemporary research, the idea reverberates in the use of predictive processing frameworks for explaining how we perceive [2,18,19] and explore natural scenes [20] as well as how they are analysed in the brain [21,22].

Together, the currently favoured theoretical frameworks converge towards a view in which the contents of our internal models shape how we perceive the world. Yet, critical questions remain unsolved: what exactly are the contents of the internal models that guide natural vision? And, given the variability in visual experiences, how do these models differ across individuals? Here, we discuss how internal models have classically been characterized in the domain of natural scene perception. Reviewing this literature, we distil a need for complementary methods that enable individual participants to report the characteristics of their internal model of a scene. We highlight emerging methods suitable for this purpose, including drawings, scene arrangements, linguistic descriptions and neuroimaging-based methods.

2. The classical approach to characterizing internal models

(a) Probing internal models of scenes by manipulating input characteristics

Previous work was built on the assumption that the contents of internal models can be studied by varying the level of typicality (i.e. schema-congruence) in the input. Under this assumption, internal models can be characterized by studying responses to stimuli that are in accordance with or in conflict with typical real-world experience, and thus with internal models (figure 1). Through such manipulations, researchers were able to isolate various aspects of typical scene structure that facilitate perception and cortical processing. These include the positioning of individual objects across space [23,27,28], spatial relationships between objects [24,29–32], contextual relationships between scenes and objects [33–36] and the spatial configuration of the scene as a whole [14,37–39]. Together, these studies show that typical scene structure at multiple levels of description contributes to the efficient perception and neural representation of scenes (for reviews, see [5–10,19]), suggesting that internal models contain rich information about the typical properties of natural scenes.

(b) Challenges for the stimulus manipulation approach

While the approach of manipulating stimulus characteristics has led to significant advances in our understanding of the contents of our internal models of the world, this ‘classical’ approach has several critical downsides.

First, it largely rests on the experimenter’s intuition of what a typical scene looks like and which factors construe its typicality. Although researchers have started using computational analyses to determine typical scene properties more objectively, for example, by extracting object distributions across large scene databases [6,40,41], such approaches are still centred on the idea that the property selected by the researcher plays an important role. Properties that intuitively should be featured in our internal models because they are visually prominent (e.g. the colour of the objects contained in a bathroom versus living room) can, in principle, be relatively uncritical for scene processing, while others that are harder to grasp intuitively (e.g. the relative orientation of objects towards each other) may be exploited by the brain more strongly.

Second, stimulus manipulation approaches only allow for independently manipulating particular stimulus dimensions at a time. Natural scenes, however, cannot be easily decomposed into a few orthogonal dimensions. Furthermore, the relevant dimensions probably interact with each other: in a kitchen scene, objects like an oven or sink are typically aligned along the walls, whereas smaller objects like utensils or cups are placed on horizontal surfaces. Studies separately investigating object distributions or scene geometry may therefore miss critical interactions between these properties. On a practical end, studies that look at such different factors tend to employ different experimental paradigms, making it hard to compare their relative contributions.

Third, many studies use artificial or unusual stimuli to create ‘atypical’ scenes. For instance, in studies of scene–object congruence [25,33,34], researchers need to create incongruent conditions, in which the objects are positioned atypically: a living room lamp is shown on a street, and the streetlight in a living room. Recent behavioural work suggests that differences between congruent and incongruent conditions may indeed be driven by a ‘congruency cost’, where the unexpected, incongruent objects gain a relative processing advantage [42]. Moreover, on the cortical level, spatially distinct regions code for congruent and incongruent conditions [43]. The problem is further aggravated when the atypical conditions violate the laws of physics [24,31] or produce unnaturalistic inputs [7].

Finally, the stimulus manipulation approach neglects inter-individual differences in internal models. Yet, internal models probably differ across individuals: although a typical kitchen will probably look somewhat similar for two people, there may be critical differences, for instance, in the placement of objects. Such individual differences may relate to idiosyncratic visual diets, but also to cultural, linguistic or socioeconomic factors [44,45]. If we could harness this variability, we may find that there are characteristic differences in the way each of us perceives the world, based on our idiosyncratic priors of what the world looks like.

To overcome these critical differences, an approach focused on obtaining descriptions of internal models directly from the participants bears enormous potential. Moving on, it is important to note that the following approaches are not intended to replace the classical approach discussed above—they rather constitute an addition to the existing toolkit for characterizing natural vision.

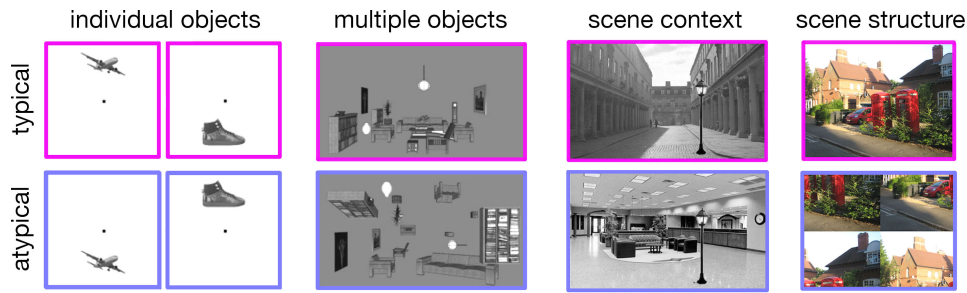


Figure 1. Understanding internal models through stimulus manipulation. To infer the contents of internal models of the world, researchers have manipulated the real-world typicality of visual inputs on different levels. From left to right: manipulations in the typical positioning of individual objects across visual space [23], the typical composition of multiple objects across space (figure is reproduced from [24]), the semantic consistency between scenes and the objects they contain [25] and the structural coherence of the scene [26]. By comparing typically arranged stimuli with atypically arranged stimuli, such studies show that the visual system preferentially processes stimuli that are in accordance with our priors.

3. A complementary approach for characterizing internal models of the world

(a) Describing internal models

Here, we highlight a novel, complementary approach that characterizes individual participants' internal models of scenes in more unconstrained ways, without prior assumptions about their properties (figure 2). The key idea is to ask participants to provide 'descriptions' about the contents of their internal model that can then inform further investigation. By obtaining such descriptions of internal models, we can start to understand how individuals converge and diverge in their conceptions of scene typicality. Moreover, we can in turn use descriptions of internal models to make targeted predictions about processing efficiency for individual scenes in individual observers.

This approach circumvents critical limitations of the classical stimulus manipulation approach. First, researchers do not need to use their intuitions about which features constrain internal models of natural scenes, as they can rather rely on the features emerging from participants' descriptions. For instance, rather than manipulating specific features based on prior knowledge, such as the object position or overall scene context (figure 1), individual descriptions provide direct access to the features prioritized across participants. Second, multiple interacting feature dimensions can be studied at once, as these will inherently be present in the descriptions. Third, there is no need to artificially create atypical stimuli: the individual typicality of a range of stimuli can be quantified relative to the descriptors. Finally, by focusing on descriptions of internal models in individuals, we can understand to what extent our priors are general or idiosyncratic.

How can we experimentally obtain such descriptions of internal models? We will discuss several methods that enable individual participants to convey their idea of a typical scene exemplar. One strong candidate is drawing, which has proven to be a versatile tool for transforming mental representations into visible descriptions [46]. Beyond that, we discuss scene arrangement, linguistic descriptions and neuroimaging-based techniques. We illustrate these techniques with a range of studies, primarily focusing on the memory and perception literature.

(b) Line drawings as descriptors of internal models

Line drawings can be seen as functional abstractions of the ways in which we see the world in the sense that they 'exploit the underlying neural codes of vision' [47]: when we draw an object or a scene, our drawing tends to focus on what conveys the most essential details of visual images in a form that abstracts away from irrelevant detail. Line drawings of scenes are recognized with virtually identical efficiency as scene photographs [48], probably because they preserve critical information about the curvature and intersection of contours [49,50]. Neuroimaging work has shown that line drawings yield characteristic category-specific neural activation patterns in the high-level visual cortex [50,51]. Beyond theoretical reasons, drawings also offer practical advantages: they are easy to generate and enable the creation of rich scenes in an (almost) unconstrained manner. These qualities render drawings an ideal candidate for the description of internal models. In the following, we highlight how these advantages have been leveraged in clinical and developmental settings, before turning towards research in memory and perception.

(i) Line drawings in clinical assessment and developmental research

The use of drawings has a long history in clinical assessment to diagnose and classify visual impairments such as agnosia [52], spatial neglect [53] or different types of neurodegenerative disease [54,55]. Furthermore, systematic differences in drawing tasks have been described for individuals with psychiatric disorders, including autism spectrum disorder [56] and schizophrenia [57]. These two conditions are associated with compromised predictive processing [58,59], suggesting that alterations to drawings in these disorders may be linked to alterations in internal models of the world.

In a similar vein, drawings have proven extremely useful for studying internal models across development. For instance, drawings have been used to assess the emergence of detail in visual representations of objects [60–62]. In recent work, Long *et al.* [61] asked children across different ages to draw various everyday objects, revealing characteristic changes in drawings across

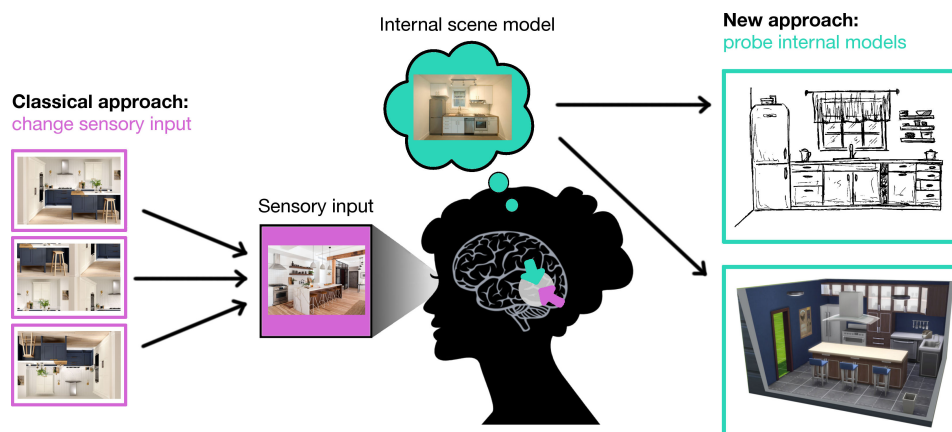


Figure 2. A complementary approach for studying internal models of the world. The classical approach aims at discovering properties of internal models through stimulus manipulation, for instance by manipulating a scene's global structure. Here, we highlight a complementary approach, in which the contents of internal models are described by observers, for instance through line drawing (where people draw typical versions of scenes) or scene arrangement methods (where people arrange physical or virtual scenes in typical ways). These descriptions can, in turn, be used to derive targeted predictions about processing efficiency for a set of inputs.

development (figure 3a). Interestingly, the content of children's drawings predicted their ability to recognize visual objects [62], suggesting a link between the mental representation revealed by drawings and the ability of the visual system to process critical object details. Together, these findings indicate that drawings mirror the visual system's ability to efficiently represent visual inputs.

(ii) Line drawings in studying memory

The use of drawings for studying internal representations in healthy adults dates back around a century. Metzger [66] described how participants made systematic drawing errors when copying line drawings under unfavourable conditions such as low contrast, miniature size or peripheral presentation. These reproductions tended to appear more unitary, regular and tightly structured than the originals, suggesting a normalization process towards internal perceptual templates.

More recently, line drawings received renewed attention in the memory literature, enabling unconstrained free-recall tasks on complex scenes [46]. For instance, Bainbridge *et al.* [67] uncovered a surprising amount of detail in memory drawings of scenes and objects. The precision of memory drawings varies as a function of scene typicality: semantically inconsistent objects are remembered more vividly than semantically consistent objects—at the expense of weaker memory for other scene characteristics in the incongruent images [63] (figure 3b). Memory drawings also reveal the degree of boundary extension in natural scene images. In a large-scale analysis, Bainbridge & Baker [64] demonstrated that this effect is ultimately stimulus-dependent, with the tendency for boundaries to be extended for scenes perceived as near and compressed for those perceived as distant (figure 4c), interpreted as a normalization towards a typical viewing distance. In this trade-off, the depth of field plays a critical role, such that a naturalistic depth of field leads to a larger boundary extension than a non-naturalistic depth of field [69]. Together, these findings further illustrate how drawings can help to understand the intricacies of human memory representations.

(iii) Line drawings in visual perception

In the study of perception, drawings have been used more sparingly. Yet, they are a promising method for studying predictive processes. In predictive processing frameworks, perception is often described as a generative model in which a percept is constructed by employing the internal world model to infer the most likely cause of sensory input [1]. Drawings offer researchers a tool to access these hypotheses by serving as an extension of the generative process into a visible format.

In line with this idea, drawings have been used to capture representations of invisible but predictable visual content in the visual cortex [65]. In this study, participants viewed natural scene images, in which one quadrant was occluded while functional magnetic resonance imaging (fMRI) activity was recorded from areas of the early visual cortex that exclusively respond to input from the occluded quadrant. Outside the scanner, participants were asked to draw the likely content of the occluded quadrant. Results showed that basic visual features of the drawings predicted scene-specific activations in the unstimulated area of the primary visual cortex (figure 4d). This finding highlights that drawings can be used to capture the content of neural predictions.

More recently, Wang *et al.* [68] used drawings as a readout of individual participants' internal models of visual scenes (figure 4). Here, participants were asked to draw typical versions of a set of natural scene categories (e.g. kitchens or living rooms). These drawings were converted into standardized three-dimensional renders to control for different drawing abilities and styles. In a subsequent categorization task, participants were more accurate in categorizing renders that were constructed from their *own* drawings (and were thus more similar to their *own* internal models) than in categorizing renders based on *other* participants' drawings (which were more dissimilar to their *own* internal models). The authors further showed that the similarity to the scene renders based on participants' own drawings (measured by a deep neural network model) predicted categorization accuracy on other rendered scenes. This result demonstrates how drawings can be used to make personalized predictions about the efficiency of perception—derived from only a single drawing of a typical scene. Complementary EEG

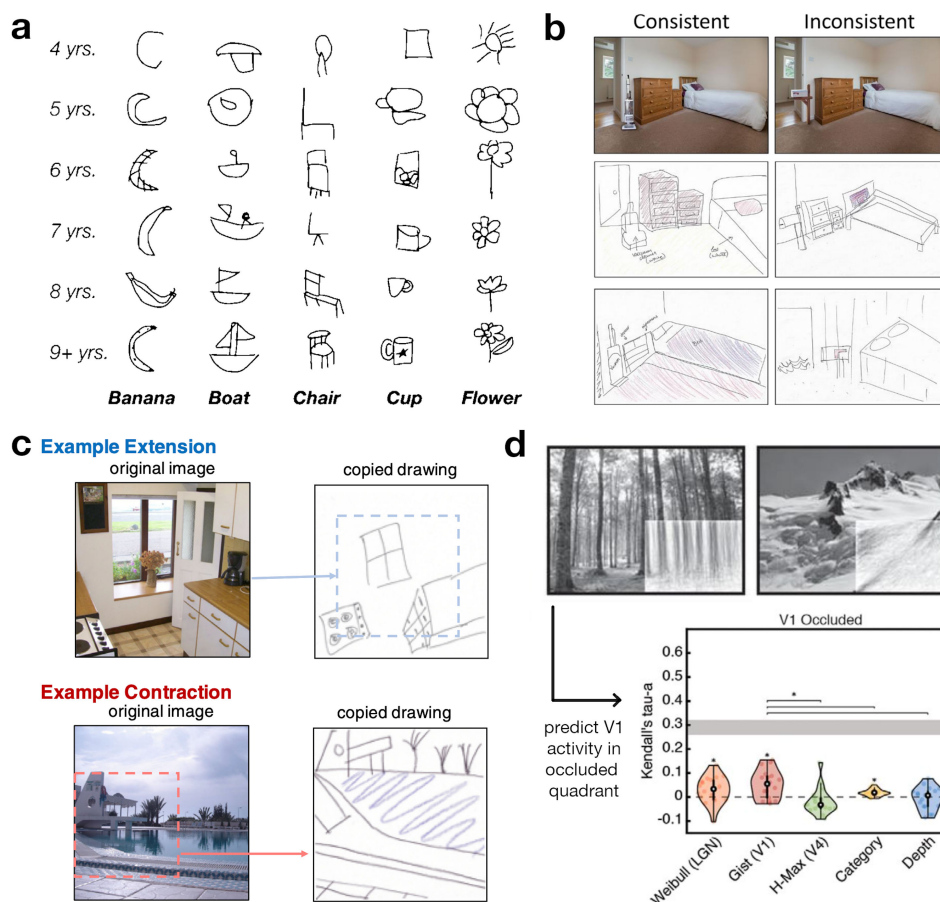


Figure 3. Using drawings to describe representations in development, memory and perception. (a) In developmental research, the use of drawings allows researchers to gain insights into the emergence of detailed visual object representations (figure is reproduced from [62] under a CC BY 4.0 license). (b) In memory research, drawings can be used to quantify memory precision in free recall paradigms. For instance, in scenes with inconsistent objects, more detail about the inconsistent object is recalled, at the expense of recalling details of the scene (figure reproduced from [63]). (c) Using a similar free recall paradigm, Bainbridge & Baker [64] showed that scene boundaries are extended or compressed in memory, depending on the viewpoint and geometry of the original scene. (d) In perception research, drawings were used to probe the cortical filling-in of missing information. Participants' drawings of what should be present in an occluded quadrant predict neural activation: response patterns in areas of primary visual cortex (V1) that respond to the occluded quadrant are well explained by visual low-level visual features of these drawings [65].

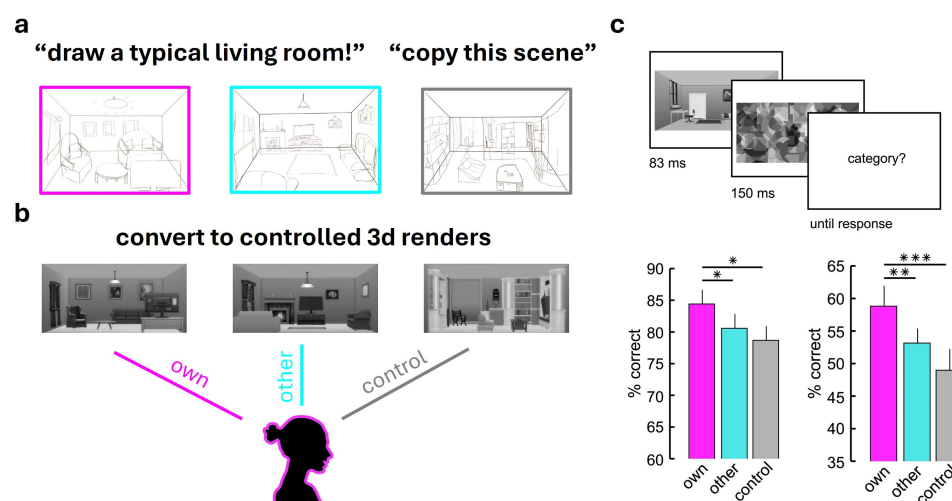


Figure 4. Using drawings to link individual differences in internal models to idiosyncrasies in perception. (a) To assess the contents of internal models for real-world scenes, participants drew typical versions of scene categories (here: living rooms). (b) These drawings were converted to three-dimensional renders to control for visual differences. (c) During the subsequent categorization task, participants categorized briefly presented renders. Critically, they viewed renders based on their own drawings ('own' condition), other participants' drawings ('other' condition) or renders created from scenes participants previously copied from a photograph ('control' condition, designed to control for drawing-related familiarity effects). Across two experiments with two (left) or six (right) scene categories, participants more accurately categorized renders from the 'own' condition than renders from the 'other' or 'control' conditions, suggesting that similarity to internal models on the individual level modulates scene processing in idiosyncratic ways [68].

work [70] showed that neural representations of scenes that are similar to participants' drawings (and thus their internal models) are enhanced during perceptual processing unfolding in the initial 250 ms of visual analysis. This suggests a rapid interaction between the visual input and an observer's internal model during natural vision, underscoring the importance of idiosyncratic representations in scene perception. Although the current results are based on a recognition task, we believe that other behaviours are also shaped by an individual's internal models. This opens new avenues for research into inter-individual variability in other tasks, for example, how individuals navigate through a scene or search for an object.

However, drawing methods also come with limitations, such as the challenge of objectively quantifying the contents of drawings and handling the substantial inter-subject variability in drawing abilities and style. Variation in drawing expertise has been associated with inter-individual differences in cognitive and perceptual abilities such as visual imagery, shape encoding and detection, as well as the allocation of visual attention and working memory [71–74]. Different approaches to minimizing such potentially confounding factors have been put forward: sufficient sample sizes and suitable control conditions can mitigate variance related to drawing abilities. Further, innovative methods like pen-tracking, computer vision or online crowdsourcing provide objective and reproducible tools for quantifying drawings [46,75].

In sum, drawings have proven a powerful tool for describing internal representations and have advanced our understanding of the precision of visual memory, the development of visual object representation and the perception and neural representations of scenes. Nevertheless, the methodology of drawings comes with limitations. In the following, we discuss three alternative methods for characterizing the contents of internal models, which offer complementary strengths.

(c) Alternative descriptors of internal models

We highlight three alternative ways of assessing the contents of internal models. The first two approaches (scene arrangement and linguistic descriptions) build on a similar principle as drawings by enabling participants to describe their internal model using physical or virtual objects, as well as language. Finally, we highlight an approach that attempts to directly infer characteristics of subjects' internal models from brain recordings without relying on overt reports.

(i) Scene arrangement

In scene arrangement paradigms, participants create a scene by arranging a set of candidate objects provided by the experimenter. Although such methods often limit participants' degrees of freedom in describing their internal models (e.g. because of fixed object exemplars available for arrangement), they mitigate the issue of inter-individual variability in drawings.

Scene arrangement tasks can be realized with real physical objects. For example, Öhlschläger & Vö [76] used an arrangement task in a doll house to test how children across different age groups honour semantic relationships (e.g. chairs and tables appear together in a dining room) or spatial regularities (e.g. chairs face the dining table) between objects (figure 5a). Results showed that children as young as 3 years respected semantic relationships but not spatial regularities among multiple related objects, while children over 4 years arranged the objects in semantically and spatially congruent ways. Using the same task, Bahn *et al.* [79] showed that the performance of scene arrangement tasks covaries with language development, suggesting a link between the linguistic concept organization and the visual rules that structure natural scenes.

Such real-world scene arrangement tasks, however, require real objects that need to be moved around in physical space. As an alternative, virtual reality (VR) allows for highly controlled, easily manipulable and interactive environments [80]. Showcasing this potential, Draschkow & Vö [78] asked participants to construct scenes that concurred with their internal models of typical scenes (e.g. placing the objects in a kitchen in a typical fashion) or that violated them (e.g. placing the same objects in an atypical fashion) (figure 5b). In subsequent experiments, participants more successfully searched and memorized scenes arranged in typical ways, compared with scenes arranged in atypical ways. These results show how descriptors of internal models, as captured by explicit scene arrangement, can in turn be used to test perception and memory in environments that are specifically tailored to individual participants' internal models.

Another promising variant of scene arrangement methods is adjustment procedures, where participants linearly vary scenes along one or more dimensions to match their internal model. Such approaches have, for instance, been used for matching colour expectations [81] or peripheral appearance [82]. For scenes, such adjustments can easily be realized for low-level properties such as contrast or colour. Yet, high-level properties can also be adjusted through continuous scene spaces derived from generative computational models, as for instance implemented in the 'scene wheel' [83], which provides seamless continua between individual scene exemplars. Such adjustment methods, however, depend on the experimenter's choice of dimensions, but may strike a balance between experimental control and participant-driven insights into internal model properties.

(ii) Linguistic descriptions

Focusing more directly on conceptual rather than visual attributes of scenes, linguistic descriptions of scenes offer another effective tool that is independent of sensory-motor demands. Such reports are comparably easy to obtain, and they can carry rich semantic detail. Yet, they may be less precise in capturing some of the aspects that drawings convey (e.g. spatial organization).

When asked to describe scene images, our linguistic descriptions are tightly linked to our perception of scenes. For example, atypical scenes are harder to describe after brief exposure compared with typical scenes [84]. Furthermore, linguistic descriptions are constrained by individual differences in the ways we explore scenes [85]: inter-subject similarities in fixation patterns during free viewing can be predicted by inter-subject similarities in subsequent scene descriptions. For instance, participants

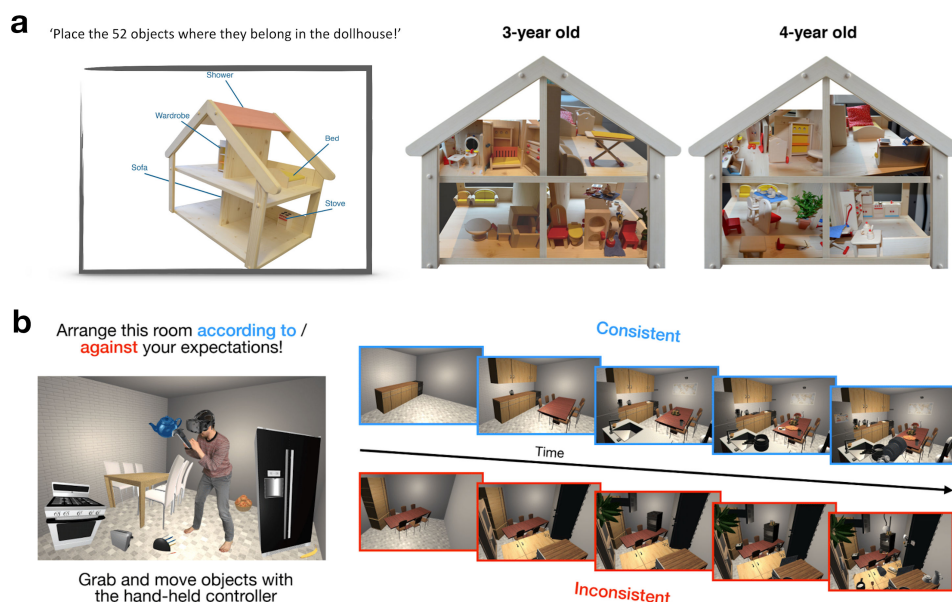


Figure 5. Using explicit scene arrangement to describe internal models. (a) Children of different age groups were asked to arrange a set of miniature objects across a dollhouse. Object arrangements showed that children first appreciate semantic object similarities and only later incorporate the typical spatial organization across groups of objects [77]. (b) Participants arranged objects in a VR environment into typical or atypical configurations. In subsequent search and memory tasks, participants performed better when the task was situated in the scenes constructed in a typical fashion (figure reproduced from [78] under a CC BY 4.0 license).

who mentioned people more often in their descriptions also looked at people more prominently during exploration, and participants who mentioned text more often spent more time looking at text. This finding highlights the potential of linguistic descriptors to capture information about scene representations on the individual level.

Future studies could employ linguistic descriptions to gauge the contents of internal models directly, revealing the conceptual factors that organize our priors. Specifically, similar to the paradigm used by Wang *et al.* [68] wherein participants were asked to draw typical versions of a set of natural scene categories and then tested on categorization for scenes similar or dissimilar to these drawings, participants could alternatively provide linguistic descriptions of what they think a typical exemplar of a specific scene category should look like. The similarity of any given scene image with the linguistic descriptions provided by individual participants could, in turn, be used to predict perceptual efficiency or neural responses for these scenes on the individual level. Alternatively, generative text-to-image models could be used to generate stimulus materials that are in accordance with, or deviate from, individual participants' linguistic descriptions. However, the outputs of generative models are influenced by the model's own priors acquired throughout training. The model's priors may be difficult to disentangle from the participant's internal model.

(iii) Neural quantification of internal models

As an alternative to behavioural reports, the content of participants' internal models could, in principle, be read out from neural responses. This is an exciting future prospect because brain activity recorded during simple visual tasks or passive fixation is potentially less prone to subjective biases and task-specific demand characteristics.

Studies linking scene typicality to neural responses are the foundation for this idea. Scenes that align with internal models strongly produce more diagnostic, 'sharpened' neural responses [86]. This variation in visual responses can be used to derive 'neural prototypes' of a given category in individual participants, for instance, by averaging responses to a larger set of exemplars into a prototypical response [87] or by comparing an exemplar's activity pattern to all other members of that category, following the premise that a typical exemplar shares more attributes with other members of its category than atypical exemplars [88]. However, more work is needed to characterize how typicality is processed in the brain before neuroimaging-based methods can be used to read out a person's internal model with confidence.

Moreover, neuroimaging-based approaches remain inherently constrained by the experimenter's selection of stimulus materials, which may not cover natural visual distributions sufficiently and in unbiased ways. To mitigate this constraint, large and diverse stimulus sets that feature thousands of scene instances, such as the Natural Scenes Dataset (NSD; [89]), provide useful benchmarks. A promising direction for future research is to combine neural measures of typicality with free production approaches, such as drawings or scene arrangement, thereby leveraging their complementary strengths.

4. Challenges in describing internal models

The methods described above enable researchers to infer the content of internal models with fewer constraints and prior assumptions about their properties than classical approaches. However, these methods also come with their own challenges. We highlight three challenges below and outline what we gain from solving them.

First, behavioural descriptions of internal models, such as drawings or linguistic reports, are subjective reports. Can we assume that such introspective insights are reliable? Introspection is often disregarded as inherently problematic because

observers may not be able to reliably characterize their internal representations [90,91]. However, this view has been challenged, most prominently by Gestalt psychologists [92], but also more recently [93,94], with proponents arguing that introspection offers converging information compared with analytic approaches. Further, the quality of introspective insight can be addressed empirically by quantifying whether introspective insights about internal models indeed predict the efficiency of perception. Our review, therefore, does not make the case to replace or overcome classical approaches to studying scene vision—we rather need to combine stimulus manipulation approaches with approaches for describing internal models.

Second, some of the outlined methods require separating informative differences from incidental variance across individuals. This is particularly relevant for drawings, where different drawing styles and abilities [95,96] introduce substantial variation that is not directly related to our internal representations. Moreover, methods like scene arrangement may bias the provided descriptions by offering a limited number of available objects. A careful choice of methods, control conditions, sample size and analysis approach is required to minimize these shortcomings and unfold the complementary strengths of these techniques. Additionally, participants may prioritize some objects over others during production. Reasons for this include biases towards larger and more salient objects (e.g. a sink is more often drawn than a toothbrush) or more scene-diagnostic objects (e.g. a socket is rarely drawn, given it is not diagnostic for a room category). Such biases may yield imperfect descriptors of internal models.

Third, although subjective reports such as drawings offer rich insights into internal models, they may not be ideal to clarify the complexity of features in the internal model. Recent work shows that individual differences in internal models are expressed in complex high-level features coded in late layers of deep neural network models [70], suggesting that complex object- and scene-related features structure our internal models. Yet, low-level properties may prominently feature in internal models, too, and drawings can capture some of such low-level regularities, such as visual density, texture continuity or edge alignment. Other low-level properties, like colour distributions or texture properties, are harder to capture with drawings and may elude investigation with drawing methods.

In addition, most of the methods highlighted here yield single descriptions of participants' internal models, implying that there is a single, stable internal model for a given scene category. However, internal models may encompass multiple typical scene configurations (e.g. private versus public bathrooms). They may also be shaped by context (e.g. in the form of precision weighting [1]), recently formed associations (e.g. through serial dependence effects [97]) and behavioural goals [98]. This requires experiments that repeatedly quantify the contents of internal models within the same participants, thereby characterizing how internal models change across time and context.

5. Conclusion

We showed that classical approaches to characterizing internal models of natural scenes through stimulus manipulation have notable limitations, including an over-reliance on *a priori* assumptions about scene typicality, restricted possibilities for stimulus manipulation, the use of artificial control stimuli and insensitivity to inter-individual differences. We therefore highlight a complementary methodological framework that allows for more unconstrained descriptors of participants' internal models. One promising method is the use of line drawings, which has recently opened new avenues in the study of visual memory and perception. Overall, we believe that natural vision research greatly benefits from methods with fewer constraints and prior assumptions about the nature of internal models, complementing traditional approaches. Although we focused on characterizing internal models for natural scenes, we believe that approaches targeting the contents of internal models can be very informative in other visual domains. Indeed, drawings have been used for understanding mental generalization in object recognition [99,100], perceptual distortions in peripheral vision [101] and individual differences in mental imagery [102]. Embracing this approach could yield novel insights into how internal models differently shape perception across individuals and across cultural and linguistic contexts, and how alterations of internal models drive changes in visual processing across the lifespan and from health to disease.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. M.E.: conceptualization, writing—original draft, writing—review and editing; S.A.: conceptualization, writing—original draft, writing—review and editing; I.D.: conceptualization, writing—original draft, writing—review and editing; G.W.: writing—original draft, writing—review and editing; M.J.F.: conceptualization, writing—original draft, writing—review and editing; R.M.C.: supervision, writing—review and editing; D.P.: supervision, writing—review and editing; D.K.: conceptualization, funding acquisition, project administration, supervision, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy (EXC 3066/1 'The Adaptive Mind', project no. 533717223) and by an ERC Starting grant (PEP, ERC-2022-STG 101076057; awarded to D.K.). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgements. We thank Melissa Vö for providing materials for figure 4.

References

1. Clark A. 2013 Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204. (doi:10.1017/S0140525X12000477)

2. de Lange FP, Heilbron M, Kok P. 2018 How do expectations shape perception? *Trends Cogn. Sci.* **22**, 764–779. (doi:10.1016/j.tics.2018.06.002)
3. Kayser C, Körding KP, König P. 2004 Processing of complex stimuli and natural scenes in the visual cortex. *Curr. Opin. Neurobiol.* **14**, 468–473. (doi:10.1016/j.conb.2004.06.002)
4. Mirza MB, Adams RA, Mathys CD, Friston KJ. 2016 Scene construction, visual foraging, and active inference. *Front. Comput. Neurosci.* **10**, 56. (doi:10.3389/fncom.2016.00056)
5. Bar M. 2004 Visual objects in context. *Nat. Rev. Neurosci.* **5**, 617–629. (doi:10.1038/nrn1476)
6. Kaiser D, Quek GL, Cichy RM, Peelen MV. 2019 Object vision in a structured world. *Trends Cogn. Sci.* **23**, 672–685. (doi:10.1016/j.tics.2019.04.013)
7. Kaiser D, Cichy RM. 2021 Parts and wholes in scene processing. *J. Cogn. Neurosci.* **34**, 4–15. (doi:10.1162/jocn_a_01788)
8. Oliva A, Torralba A. 2007 The role of context in object recognition. *Trends Cogn. Sci.* **11**, 520–527. (doi:10.1016/j.tics.2007.09.009)
9. Võ MH. 2021 The meaning and structure of scenes. *Vision Res.* **181**, 10–20. (doi:10.1016/j.visres.2020.11.003)
10. Võ MH, Boettcher SE, Draschkow D. 2019 Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Curr. Opin. Psychol.* **29**, 205–210. (doi:10.1016/j.copsyc.2019.03.009)
11. Von Helmholtz H. 1867 *Treatise on physiological optics*. vol. III. Garden City, New York, USA: Dover Publications.
12. Mandler JM. 1984 *Stories, scripts, and scenes: aspects of schema theory*. New York, NY, USA: Psychology Press. (doi:10.4324/9781315802459)
13. Minsky M. 1974 *A framework for representing knowledge*. Cambridge, MA, USA: MIT. See <https://dspace.mit.edu/bitstream/handle/1721.1/6089/AIM-306.pdf?%2520sequence%3D2>.
14. Biederman I. 1972 Perceiving real-world scenes. *Science* **177**, 77–80. (doi:10.1126/science.177.4043.77)
15. Biederman I, Mezzanotte RJ, Rabinowitz JC. 1982 Scene perception: detecting and judging objects undergoing relational violations. *Cognit. Psychol.* **14**, 143–177. (doi:10.1016/0010-0285(82)90007-X)
16. Brewer WF, Treysen JC. 1981 Role of schemata in memory for places. *Cognit. Psychol.* **13**, 207–230. (doi:10.1016/0010-0285(81)90008-6)
17. Mandler JM, Parker RE. 1976 Memory for descriptive and spatial information in complex pictures. *J. Exp. Psychol.* **38**–48. (doi:10.1037/0278-7393.2.1.38)
18. Bar M. 2009 The proactive brain: memory for predictions. *Phil. Trans. R. Soc. B* **364**, 1235–1243. (doi:10.1098/rstb.2008.0310)
19. Peelen MV, Berlot E, de Lange FP. 2024 Predictive processing of scenes and objects. *Nat. Rev. Psychol.* **3**, 13–26. (doi:10.1038/s44159-023-00254-0)
20. Henderson JM. 2017 Gaze control as prediction. *Trends Cogn. Sci.* **21**, 15–23. (doi:10.1016/j.tics.2016.11.003)
21. Kaiser D, Jacopo T, Cichy RM. 2019 A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *eLife* **8**. (doi:10.7554/eLife.48182)
22. Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E. 2015 Contextual feedback to superficial layers of V1. *Curr. Biol.* **25**, 2690–2695. (doi:10.1016/j.cub.2015.08.057)
23. Kaiser D, Cichy RM. 2018 Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *J. Neurophysiol.* **120**, 848–853. (doi:10.1152/jn.00229.2018)
24. Bilalić M, Lindig T, Turella L. 2019 Parsing rooms: the role of the PPA and RSC in perceiving object relations and spatial layout. *Brain Struct. Funct.* **224**, 2505–2524.
25. Munneke J, Brentari V, Peelen M. 2013 The influence of scene context on object recognition is independent of attentional focus. *Front. Psychol.* **4**, 1–10. (doi:10.3389/fpsyg.2013.00552)
26. Kaiser D, Häberle G, Cichy RM. 2021 Coherent natural scene structure facilitates the extraction of task-relevant object information in visual cortex. *Neuroimage* **240**, 118365. (doi:10.1016/j.neuroimage.2021.118365)
27. Kaiser D, Moeskops MM, Cichy RM. 2018 Typical retinotopic locations impact the time course of object coding. *Neuroimage* **176**, 372–379. (doi:10.1016/j.neuroimage.2018.05.006)
28. Kaiser D, Cichy RM. 2018 Typical visual-field locations facilitate access to awareness for everyday objects. *Cognition* **180**, 118–122. (doi:10.1016/j.cognition.2018.07.009)
29. Gronau N, Shachar M. 2014 Contextual integration of visual objects necessitates attention. *Atten. Percept. Psychophys.* **76**, 695–714. (doi:10.3758/s13414-013-0617-8)
30. Kaiser D, Stein T, Peelen MV. 2014 Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc. Natl Acad. Sci. USA* **111**, 11217–11222. (doi:10.1073/pnas.1400559111)
31. Kaiser D, Peelen MV. 2018 Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *Neuroimage* **169**, 334–341. (doi:10.1016/j.neuroimage.2017.12.065)
32. Kim JG, Biederman I. 2011 Where do objects become scenes? *Cereb. Cortex* **21**, 1738–1746. (doi:10.1093/cercor/bhq240)
33. Chen L, Cichy RM, Kaiser D. 2022 Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. *Cereb. Cortex* **32**, 3553–3567. (doi:10.1093/cercor/bhab433)
34. Davenport JL, Potter MC. 2004 Scene consistency in object and background perception. *Psychol. Sci.* **15**, 559–564. (doi:10.1111/j.0956-7976.2004.00719.x)
35. Mudrik L, Lamy D, Deouell LY. 2010 ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* **48**, 507–517. (doi:10.1016/j.neuropsychologia.2009.10.011)
36. Võ MH, Wolfe JM. 2013 Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol. Sci.* **24**, 1816–1823. (doi:10.1177/0956797613476955)
37. Biederman I, Rabinowitz JC, Glass AL, Stacy EW. 1974 On the information extracted from a glance at a scene. *J. Exp. Psychol.* **103**, 597–600. (doi:10.1037/h0037158)
38. Kaiser D, Häberle G, Cichy RM. 2020 Cortical sensitivity to natural scene structure. *Hum. Brain Mapp.* **41**, 1286–1295. (doi:10.1002/hbm.24875)
39. Kaiser D, Häberle G, Cichy RM. 2020 Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *J. Neurophysiol.* **124**, 145–151. (doi:10.1152/jn.00164.2020)
40. Bonner MF, Epstein RA. 2021 Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nat. Commun.* **12**, 4081. (doi:10.1038/s41467-021-24368-2)
41. Gregorová K, Turini J, Gagl B, Võ MH. 2023 Access to meaning from visual input: object and word frequency effects in categorization behavior. *J. Exp. Psychol. Gen.* **152**, 2861–2881. (doi:10.1037/xge0001342)
42. Spaak E, Peelen MV, de Lange FP. 2022 Scene context impairs perception of semantically congruent objects. *Psychol. Sci.* **33**, 299–313. (doi:10.1177/09567976211032676)
43. Faivre N, Dubois J, Schwartz N, Mudrik L. 2019 Imaging object–scene relations processing in visible and invisible natural scenes. *Sci. Rep.* **9**, 4567. (doi:10.1038/s41598-019-38654-z)
44. Barrett HC. 2020 Towards a cognitive science of the human: cross-cultural approaches and their urgency. *Trends Cogn. Sci.* **24**, 620–638. (doi:10.1016/j.tics.2020.05.007)
45. Hartley CA. 2022 How do natural environments shape adaptive cognition across the lifespan? *Trends Cogn. Sci.* **26**, 1029–1030. (doi:10.1016/j.tics.2022.10.002)
46. Fan JE, Bainbridge WA, Chamberlain R, Wammes JD. 2023 Drawing as a versatile cognitive tool. *Nat. Rev. Psychol.* **2**, 556–568. (doi:10.1038/s44159-023-00212-w)
47. Sayim B, Cavanagh P. 2011 What line drawings reveal about the visual brain. *Front. Hum. Neurosci.* **5**, 118. (doi:10.3389/fnhum.2011.00118)
48. Biederman I, Ju G. 1988 Surface versus edge-based determinants of visual recognition. *Cognit. Psychol.* **20**, 38–64. (doi:10.1016/0010-0285(88)90024-2)
49. Barrow HG, Tenenbaum JM. 1981 Interpreting line drawings as three-dimensional surfaces. *Artif. Intell.* **17**, 75–116. (doi:10.1016/0004-3702(81)90021-7)

50. Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L. 2011 Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc. Natl Acad. Sci. USA* **108**, 9661–9666. (doi:10.1073/pnas.1015666108)
51. Singer JJD, Cichy RM, Hebart MN. 2023 The spatiotemporal neural dynamics of object recognition for natural images and line drawings. *J. Neurosci.* **43**, 484–500. (doi:10.1523/JNEUROSCI.1546-22.2022)
52. Bauer RM. 2006 The agnosia. In *Clinical neuropsychology: a pocket handbook for assessment* (eds PJ Snyder, PD Nussbaum, DL Robins), pp. 508–533, 2nd edn. Washington, DC, USA: American Psychological Association. (doi:10.1037/11299-020)
53. Agrell B, Dehlin O. 1998 The clock-drawing test. *Age Ageing* **27**, 399–403. (doi:10.1093/ageing/27.3.399)
54. Cahn DA, Salmon DP, Monsch AU, Butters N, Wiederholt WC, Corey-Bloom J, Barrett-Connor E. 1996 Screening for dementia of the Alzheimer type in the community: the utility of the Clock Drawing Test. *Arch. Clin. Neuropsychol. Off. J. Natl Acad. Neuropsychol.* **11**, 529–539.
55. Wechsler D. 2009 *Wechsler memory scale*, 4th edn. London, UK: Pearson. (doi:10.1037/t15175-000)
56. Shi F, Sun W, Duan H, Liu X, Hu M, Wang W, Zhai G. 2021 Drawing reveals hallmarks of children with autism. *Displays* **67**, 102000. (doi:10.1016/j.displa.2021.102000)
57. Bozikas VP, Kosmidis MH, Gamvroula K, Hatzigeorgiadou M, Kourtis A, Karavatos A. 2004 Clock drawing test in patients with schizophrenia. *Psychiatry Res.* **121**, 229–238. (doi:10.1016/j.psychres.2003.07.003)
58. Fletcher PC, Frith CD. 2009 Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* **10**, 48–58. (doi:10.1038/nrn2536)
59. Pellicano E, Burr D. 2012 When the world becomes ‘too real’: a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510. (doi:10.1016/j.tics.2012.08.009)
60. Karmiloff-Smith A. 1990 Constraints on representational change: evidence from children's drawing. *Cognition* **34**, 57–83. (doi:10.1016/0010-0277(90)90031-E)
61. Long B, Fan J, Chai Z, Frank MC. 2019 Developmental changes in the ability to draw distinctive features of object categories. Seattle, WA, USA: Cognitive Science Society. (doi:10.31234/osf.io/8rzku). <https://escholarship.org/uc/item/1gx7k7p7>.
62. Long B, Fan JE, Huey H, Chai Z, Frank MC. 2024 Parallel developmental changes in children's production and recognition of line drawings of visual concepts. *Nat. Commun.* **15**, 1191. (doi:10.1038/s41467-023-44529-9)
63. Bainbridge WA, Kwok WY, Baker CI. 2021 Disrupted object-scene semantics boost scene recall but diminish object recall in drawings from memory. *Mem. Cognit.* **49**, 1568–1582. (doi:10.3758/s13421-021-01180-3)
64. Bainbridge WA, Baker CI. 2020 Boundaries extend and contract in scene memory depending on image properties. *Curr. Biol.* **30**, 537–543. (doi:10.1016/j.cub.2019.12.004)
65. Morgan AT, Petro LS, Muckli L. 2019 Scene representations conveyed by cortical feedback to early visual cortex can be described by line drawings. *J. Neurosci.* **39**, 9410–9423. (doi:10.1523/JNEUROSCI.0852-19.2019)
66. Metzger W. 1936 *Gesetze des Sehens. [Laws of vision]*. Frankfurt a.M., Germany: Kramer.
67. Bainbridge WA, Hall EH, Baker CI. 2019 Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nat. Commun.* **10**, 5. (doi:10.1038/s41467-018-07830-6)
68. Wang G, Foxwell MJ, Cichy RM, Pitcher D, Kaiser D. 2024 Individual differences in internal models explain idiosyncrasies in scene perception. *Cognition* **245**, 105723. (doi:10.1016/j.cognition.2024.105723)
69. Gandolfo M, Nägele H, Peelen MV. 2023 Predictive processing of scene layout depends on naturalistic depth of field. *Psychol. Sci.* **34**, 394–405. (doi:10.1177/09567976221140341)
70. Wang G, Chen L, Cichy RM, Kaiser D. 2025 Enhanced and idiosyncratic neural representations of personally typical scenes. *Proc. R. Soc. B* **292**, 20250272. (doi:10.1098/rspb.2025.0272)
71. Calabrese L, Marucci FS. 2006 The influence of expertise level on the visuo-spatial ability: differences between experts and novices in imagery and drawing abilities. *Cogn. Process* **7**, 118–120. (doi:10.1007/s10339-006-0094-2)
72. Chamberlain R, Drake JE, Kozbelt A, Hickman R, Siev J, Wagemans J. 2019 Artists as experts in visual cognition: an update. *Psychol. Aesthet. Creat. Arts* **13**, 58–73. (doi:10.1037/aca0000156)
73. Kozbelt A. 2001 Artists as experts in visual cognition. *Vis. Cogn.* **8**, 705–723. (doi:10.1080/13506280042000090)
74. Perdreau F, Cavanagh P. 2015 Drawing experts have better visual memory while drawing. *J. Vis.* **15**, 5. (doi:10.1167/15.5.5)
75. Bainbridge WA. 2022 A tutorial on capturing mental representations through drawing and crowd-sourced scoring. *Behav. Res. Methods* **54**, 663–675. (doi:10.3758/s13428-021-01672-9)
76. Öhlschläger S, Vö MH. 2020 Development of scene knowledge: evidence from explicit and implicit scene knowledge measures. *J. Exp. Child Psychol.* **194**, 104782. (doi:10.1016/j.jecp.2019.104782)
77. Öhlschläger S, Vö MH. 2017 SCEGRAM: an image database for semantic and syntactic inconsistencies in scenes. *Behav. Res. Methods* **49**. (doi:10.3758/s13428-016-0820-3)
78. Draschkow D, Vö MH. 2017 Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Sci. Rep.* **7**, 16471. (doi:10.1038/s41598-017-16739-x)
79. Bahn D, Türk DD, Tsenkova N, Schwarzer G, Vö MH, Kauschke C. 2025 Processing of scene-grammar inconsistencies in children with developmental language disorder—insights from implicit and explicit measures. *Brain Sci.* **15**, 139. (doi:10.3390/brainsci15020139)
80. Wilson CJ, Soranzo A. 2015 The use of virtual reality in psychology: a case study in visual perception. *Comput. Math. Methods Med.* 151702. (doi:10.1155/2015/151702)
81. Hansen T, Olkkonen M, Walter S, Gegenfurtner KR. 2006 Memory modulates color appearance. *Nat. Neurosci.* **9**, 1367–1368. (doi:10.1038/nn1794)
82. Valsecchi M, Koenderink J, van Doorn A, Gegenfurtner KR. 2018 Prediction shapes peripheral appearance. *J. Vis.* **18**, 21. (doi:10.1167/18.13.21)
83. Son G, Walther DB, Mack ML. 2021 Scene wheels: measuring perception and memory of real-world scenes with a continuous stimulus space. *Behav. Res. Methods* **54**, 444–456. (doi:10.3758/s13428-021-01630-5)
84. Greene MR, Botros AP, Beck DM, Fei-Fei L. 2015 What you see is what you expect: rapid scene understanding benefits from prior experience. *Atten. Percept. Psychophys.* **77**, 1239–1251. (doi:10.3758/s13414-015-0859-8)
85. Kollenda D, Reher AS, de Haas B. 2025 Individual gaze predicts individual scene descriptions. *Sci. Rep.* **15**, 9443. (doi:10.1038/s41598-025-94056-4)
86. Torralba A, Walther DB, Chai B, Caddigan E, Fei-Fei L, Beck DM. 2013 Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PLoS One* **8**, e58594. (doi:10.1371/journal.pone.0058594)
87. Iordan MC, Greene MR, Beck DM, Fei-Fei L. 2016 Typicality sharpens category representations in object-selective cortex. *Neuroimage* **134**, 170–179. (doi:10.1016/j.neuroimage.2016.04.012)
88. Davis T, Poldrack RA. 2014 Quantifying the internal structure of categories using a neural typicality measure. *Cereb. Cortex* **24**, 1720–1737. (doi:10.1093/cercor/bht014)
89. Allen EJ *et al.* 2022 A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126. (doi:10.1038/s41593-021-00962-x)
90. Engelbert M, Carruthers P. 2010 Introspection. *WIREs Cogn. Sci.* **1**, 245–253. (doi:10.1002/wcs.4)

91. Schwitzgebel E. 2008 The unreliability of naive introspection. *Philos. Rev.* **117**, 245–273.
92. Koffka K. 1924 Introspection and the method of psychology. *Br. J. Psychol.* **15**, 149–161.
93. Jack AI, Roepstorff A. 2002 Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends Cogn. Sci.* **6**, 333–339. (doi:10.1016/s1364-6613(02)01941-1)
94. Locke EA. 2009 It's time we brought introspection out of the closet. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* **4**, 24–25. (doi:10.1111/j.1745-6924.2009.01090.x)
95. Chamberlain R. 2018 Drawing as a window onto expertise. *Curr. Dir. Psychol. Sci.* **27**, 501–507. (doi:10.1177/0963721418797301)
96. Chamberlain R, McManus C, Riley H, Rankin Q, Brunswick N. 2014 Cain's house task revisited and revived: extending theory and methodology for quantifying drawing accuracy. *Psychol. Aesthet. Creat. Arts* **8**, 152–167. (doi:10.1037/a0035635)
97. Fischer J, Whitney D. 2014 Serial dependence in visual perception. *Nat. Neurosci.* **17**, 738–743. (doi:10.1038/nn.3689)
98. Bracci S, op de Beeck HP. 2023 Understanding human object vision: a picture is worth a thousand representations. *Annu. Rev. Psychol.* **74**, 113–135. (doi:10.1146/annurev-psych-032720-041031)
99. Tiedemann H, Morgenstern Y, Schmidt F, Fleming RW. 2022 One-shot generalization in humans revealed through a drawing task. *eLife* **11**, e75485. (doi:10.7554/eLife.75485)
100. Schmidt F, Tiedemann H, Fleming RW, Morgenstern Y. 2025 Inferring shape transformations in a drawing task. *Mem. Cognit.* **53**, 189–199. (doi:10.3758/s13421-023-01452-0)
101. Baldwin J, Burleigh A, Pepperell R, Ruta N. 2016 The perceived size and shape of objects in peripheral vision. *Percept* **7**, 2041669516661900. (doi:10.1177/2041669516661900)
102. Bainbridge WA, Pounder Z, Eardley AF, Baker CI. 2021 Quantifying aphantasia through drawing: those without visual imagery show deficits in object but not spatial memory. *Cortex* **135**, 159–172. (doi:10.1016/j.cortex.2020.11.014)